A multivariate generalization of Costa's entropy power inequality

Miquel Payaró and Daniel P. Palomar
Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong.
{miquel.payaro, palomar}@ust.hk

***Abstract*— A simple multivariate version of Costa's entropy power inequality is proved. In particular, it is shown that if independent white Gaussian noise is added to an arbitrary multivariate signal, the entropy power of the resulting random variable is a multidimensional concave function of the individual variances of the components of the signal. As a side result, we also give an expression for the Hessian matrix of the entropy and entropy power functions with respect to the variances of the signal components, which is an interesting result in its own right.**

I. Introduction

The entropy power of the random vector $\boldsymbol{Y} \in \mathbb{R}^n$ was first introduced by Shannon in his seminal work [1] and is, since then, defined as

$$N(\boldsymbol{Y}) = \frac{1}{2\pi \mathrm{e}} \exp\left(\frac{2}{n} h(\boldsymbol{Y})\right), \tag{1}$$

where $h(\boldsymbol{Y})$ represents the differential entropy, which, for continuous random vectors reads as[1]

$$h(\boldsymbol{Y}) = \mathsf{E}\left\{\log \frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right\}.$$

For the case where the distribution of $\boldsymbol{Y}$ assigns positive mass to one or more singletons in $\mathbb{R}^n$, the above definition is extended with $h(\boldsymbol{Y}) = -\infty$.

The entropy power of a random vector $\boldsymbol{Y}$ represents the variance (or power) of a standard Gaussian random vector $\boldsymbol{Y}_G \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$ such that both $\boldsymbol{Y}$ and $\boldsymbol{Y}_G$ have identical differential entropy, $h(\boldsymbol{Y}_G) = h(\boldsymbol{Y})$.

A. Shannon's entropy power inequality (EPI)

For any two independent arbitrary random vectors $\boldsymbol{X} \in \mathbb{R}^n$ and $\boldsymbol{W} \in \mathbb{R}^n$, Shannon gave in [1] the following inequality:

$$N(\boldsymbol{X} + \boldsymbol{W}) \geq N(\boldsymbol{X}) + N(\boldsymbol{W}).$$

The first rigorous proof of Shannon's EPI was given in [2] by Stam, and was simplified by Blachman in [3]. A simple and very elegant proof by Verdú and Guo based on estimation theoretic considerations has recently appeared in [4].

Among many other important results, Bergmans' proof of the converse for the degraded Gaussian broadcast channel [5] and Oohama's partial solution to the rate distortion region problem for Gaussian multiterminal source coding systems [6] follow from Shannon's EPI.

[1]Throughout this paper we work with natural logarithms.

B. Costa's EPI

Under the setting of Shannon's EPI, Costa proved in [7] that, provided that the random vector $\boldsymbol{W}$ is white Gaussian distributed, then Shannon's EPI can be strengthened to

$$N(\boldsymbol{X} + \sqrt{t}\boldsymbol{W}) \geq (1-t)N(\boldsymbol{X}) + tN(\boldsymbol{X} + \boldsymbol{W}), \tag{2}$$

where $t \in [0, 1]$. As Costa noted, the above EPI is equivalent to the concavity of the entropy power function $N(\boldsymbol{X} + \sqrt{t}\boldsymbol{W})$ with respect to the parameter t, or, formally[2]

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} N\big(\boldsymbol{X} + \sqrt{t}\boldsymbol{W}\big) \leq 0. \tag{3}$$

Due to its inherent interest and to the fact that the proof by Costa was rather involved, simplified proofs of his result have been subsequently given in [8]–[11].

Additionally, in his paper Costa presented two extensions of his main result in (3). Precisely, he showed that the EPI is also valid when the Gaussian vector $\boldsymbol{W}$ is not white, and also for the case where the t parameter is multiplying the arbitrarily distributed random vector $\boldsymbol{X}$,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} N\big(\sqrt{t}\boldsymbol{X} + \boldsymbol{W}\big) \leq 0. \tag{4}$$

Similarly to Shannon's EPI, Costa's EPI has been used successfully to derive important information-theoretic results concerning, e.g., Gaussian interference channels in [12] or multi-antenna flat fading channels with memory in [13].

C. Aim of the paper

Our objective is to extend the particular case in (4) of Costa's EPI to the multivariate case, allowing the real parameter $t \in \mathbb{R}$ to become a matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$, which, to the best of the authors' knowledge, has not been considered before.

Beyond its theoretical interest, the motivation behind this study is due to the fact that the concavity of the entropy power with respect to $\mathbf{T}$ implies the concavity of the entropy and mutual information quantities, which would be a very desirable property in optimization procedures in order to be able to, e.g., design the linear precoder that maximizes the

[2]The equivalence between equations (2) and (3) is due to the fact that the function $N\big(\boldsymbol{X} + \sqrt{t}\boldsymbol{W}\big)$ is twice differentiable almost everywhere thanks to the smoothing properties of the added Gaussian noise.

mutual information in the linear vector Gaussian channel with arbitrary input distributions.

Consequently, we investigate the concavity of the function

$$N\big(\mathbf{T}^{1/2}\boldsymbol{X}+\boldsymbol{W}\big), \tag{5}$$

with respect to the symmetric matrix $\mathbf{T}=\mathbf{T}^{1/2}\mathbf{T}^{T/2}$. Unfortunately, the concavity in $\mathbf{T}$ of the entropy power can be easily disproved by finding simple counterexamples as in [14] or even through numerical computations of the entropy power.

Knowing this negative result, we thus focus our study on the next possible multivariate candidate: a diagonal matrix. Our objective now is to study the concavity of

$$N\big(\mathbf{\Lambda}^{1/2}\boldsymbol{X}+\boldsymbol{W}\big), \tag{6}$$

w.r.t. the diagonal matrix $\mathbf{\Lambda}=\operatorname{diag}(\boldsymbol{\lambda})$, with $[\boldsymbol{\lambda}]_i=\lambda_i$.

For the sake of notation, throughout this work we define

$$\boldsymbol{Y}=\mathbf{\Lambda}^{1/2}\boldsymbol{X}+\boldsymbol{W},$$

where we recall that the random vector $\boldsymbol{W}$ is assumed to follow a white zero-mean Gaussian distribution and the distribution of the random vector $\boldsymbol{X}$ is arbitrary. In particular, the distribution of $\boldsymbol{X}$ is allowed to assign positive mass to one or more singletons in $\mathbb{R}^n$. Consequently, the results presented in Theorems 1 and 2 in Section III also hold for the case where the random vector $\boldsymbol{X}$ is discrete.

II. Mathematical preliminaries

In this section we present a number of lemmas followed by a proposition that will prove useful in the proof of our multidimensional EPI. In our derivations, the identity matrix is denoted by $\mathbf{I}$, the vector with all its entries equal to 1 is represented by $\mathbf{1}$, and $\mathbf{A}\circ\mathbf{B}$ represents the Hadamard (or Schur) element-wise matrix product.

Lemma 1 (Bhatia [15, p. 15]): Let $\mathbf{A}\in\mathbb{S}_+^n$ be a positive semidefinite matrix, $\mathbf{A}\geq\mathbf{0}$. Then it follows that

$$\begin{bmatrix}\mathbf{A} & \mathbf{A}\\ \mathbf{A} & \mathbf{A}\end{bmatrix}\geq\mathbf{0}.$$

Proof: Since $\mathbf{A}\geq\mathbf{0}$, consider $\mathbf{A}=\mathbf{C}\mathbf{C}^T$ and write

$$\begin{bmatrix}\mathbf{A} & \mathbf{A}\\ \mathbf{A} & \mathbf{A}\end{bmatrix}=\begin{bmatrix}\mathbf{C}\\ \mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{C}^T & \mathbf{C}^T\end{bmatrix}.$$

■

Lemma 2 (Bhatia [15, Exercise 1.3.10]): Let $\mathbf{A}\in\mathbb{S}_{++}^n$ be a positive definite matrix, $\mathbf{A}>\mathbf{0}$. Then it follows that

$$\begin{bmatrix}\mathbf{A} & \mathbf{I}\\ \mathbf{I} & \mathbf{A}^{-1}\end{bmatrix}\geq\mathbf{0}. \tag{7}$$

Proof: Consider again $\mathbf{A}=\mathbf{C}\mathbf{C}^T$, then we have $\mathbf{A}^{-1}=\mathbf{C}^{-T}\mathbf{C}^{-1}$. Now, simply write (7) as

$$\begin{bmatrix}\mathbf{A} & \mathbf{I}\\ \mathbf{I} & \mathbf{A}^{-1}\end{bmatrix}=\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{-T}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{I}\\ \mathbf{I} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{C}^T & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{-1}\end{bmatrix},$$

which, from Sylvester's law of inertia for congruent matrices [15, p. 5] and Lemma 1, is positive semidefinite. ■

Lemma 3 (Schur Theorem): If the matrices $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, then so is the product $\mathbf{A}\circ\mathbf{B}$. If, both $\mathbf{A}$ and $\mathbf{B}$ are positive definite, then so is $\mathbf{A}\circ\mathbf{B}$. In other words, the class of positive (semi)definite matrices is closed under the Hadamard product.

Proof: See [16, Th. 7.5.3] or [17, Th. 5.2.1]. ■

Lemma 4 (Schur complement): Let the matrices $\mathbf{A}\in\mathbb{S}_{++}^n$ and $\mathbf{B}\in\mathbb{S}_{++}^m$ be positive definite, $\mathbf{A}>\mathbf{0}$ and $\mathbf{B}>\mathbf{0}$, and not necessarily of the same dimension. Then the following statements are equivalent

1) $\begin{bmatrix}\mathbf{A} & \mathbf{D}\\ \mathbf{D}^T & \mathbf{B}\end{bmatrix}\geq 0$,
2) $\mathbf{B}\geq\mathbf{D}^T\mathbf{A}^{-1}\mathbf{D}$,
3) $\mathbf{A}\geq\mathbf{D}\mathbf{B}^{-1}\mathbf{D}^T$,

where $\mathbf{D}\in\mathbb{R}^{n\times m}$ is any arbitrary matrix.

Proof: See [16, Th. 7.7.6] and the second exercise following it or [18, Prop. 8.2.3]. ■

With the above lemmas at hand, we are now ready to prove the following proposition:

Proposition 5: Consider two positive definite matrices $\mathbf{A}\in\mathbb{S}_{++}^n$ and $\mathbf{B}\in\mathbb{S}_{++}^n$ of the same dimension, and let $\mathbf{D}_\mathbf{A}$ be a diagonal matrix containing the diagonal elements of $\mathbf{A}$, (i.e., $\mathbf{D}_\mathbf{A}=\mathbf{A}\circ\mathbf{I}$). Then it follows that

$$\mathbf{A}\circ\mathbf{B}^{-1}\geq\mathbf{D}_\mathbf{A}\left(\mathbf{A}\circ\mathbf{B}\right)^{-1}\mathbf{D}_\mathbf{A}. \tag{8}$$

Proof: From Lemmas 1, 2, and 3, it follows that

$$\begin{bmatrix}\mathbf{A} & \mathbf{A}\\ \mathbf{A} & \mathbf{A}\end{bmatrix}\circ\begin{bmatrix}\mathbf{B} & \mathbf{I}\\ \mathbf{I} & \mathbf{B}^{-1}\end{bmatrix}=\begin{bmatrix}\mathbf{A}\circ\mathbf{B} & \mathbf{D}_\mathbf{A}\\ \mathbf{D}_\mathbf{A} & \mathbf{A}\circ\mathbf{B}^{-1}\end{bmatrix}\geq 0.$$

Now, from Lemma 4, the result follows directly. ■

Corollary 6: Let $\mathbf{A}\in\mathbb{S}_{++}^n$ be a positive definite matrix. Then,

$$\mathbf{d}_\mathbf{A}^T\left(\mathbf{A}\circ\mathbf{A}\right)^{-1}\mathbf{d}_\mathbf{A}\leq n, \tag{9}$$

where we have defined $\mathbf{d}_\mathbf{A}=\mathbf{D}_\mathbf{A}\mathbf{1}=(\mathbf{A}\circ\mathbf{I})\mathbf{1}$ as a column vector with the diagonal elements of matrix $\mathbf{A}$.

Proof: Particularizing the result in Proposition 5 with $\mathbf{B}=\mathbf{A}$ and pre- and post-multiplying it by $\mathbf{1}^T$ and $\mathbf{1}$ we obtain

$$\mathbf{1}^T\left(\mathbf{A}\circ\mathbf{A}^{-1}\right)\mathbf{1}\geq\mathbf{1}^T\mathbf{D}_\mathbf{A}\left(\mathbf{A}\circ\mathbf{A}\right)^{-1}\mathbf{D}_\mathbf{A}\mathbf{1}.$$

The result in (9) now follows straightforwardly from the fact $\mathbf{1}^T\left(\mathbf{A}\circ\mathbf{A}^{-T}\right)\mathbf{1}=n$, [19] (see also [18, Fact 7.6.10], [17, Lemma 5.4.2(a)]). Note that $\mathbf{A}$ is symmetric and thus $\mathbf{A}^T=\mathbf{A}$ and $\mathbf{A}^{-T}=\mathbf{A}^{-1}$. ■

Remark 7: Note that the proof of Corollary 6 is based on the result of Proposition 5 in (8). An alternative proof could follow similarly from a different inequality by Styan in [20]

$$\mathbf{R}\circ\mathbf{R}^{-1}+\mathbf{I}\geq 2\left(\mathbf{R}\circ\mathbf{R}\right)^{-1},$$

where $\mathbf{R}$ is constrained to be a correlation matrix $\mathbf{R}\circ\mathbf{I}=\mathbf{I}$.

Proposition 8: Consider now the positive semidefinite matrix $\mathbf{A}\in\mathbb{S}_+^n$. Then,

$$\mathbf{A}\circ\mathbf{A}\geq\frac{\mathbf{d}_\mathbf{A}\mathbf{d}_\mathbf{A}^T}{n}.$$

Proof: For the case where $\mathbf{A} \in \mathbb{S}^n_{++}$ is positive definite, from (9) in Corollary 6 and Lemma 4, it follows that

$$\begin{bmatrix} \mathbf{A} \circ \mathbf{A} & \mathbf{d}_{\mathbf{A}} \\ \mathbf{d}_{\mathbf{A}}^T & n \end{bmatrix} \geq 0.$$

Applying again Lemma 4, we get

$$\mathbf{A} \circ \mathbf{A} \geq \frac{\mathbf{d}_{\mathbf{A}} \mathbf{d}_{\mathbf{A}}^T}{n}. \quad (10)$$

Now, assume that $\mathbf{A} \in \mathbb{S}^n_+$ is positive semidefinite. We thus define $\epsilon > 0$ and consider the positive definite matrix $\mathbf{A} + \epsilon \mathbf{I}$. From (10), we know that

$$(\mathbf{A} + \epsilon \mathbf{I}) \circ (\mathbf{A} + \epsilon \mathbf{I}) \geq \frac{\mathbf{d}_{\mathbf{A}+\epsilon\mathbf{I}} \mathbf{d}_{\mathbf{A}+\epsilon\mathbf{I}}^T}{n}.$$

Taking the limit as ϵ tends to 0, from continuity, the validity of (10) for positive semidefinite matrices follows. ∎

Finally, to end this section about mathematical preliminaries, we give a very brief overview on some basic definitions related to minimum mean-square error (MMSE) estimation. These definitions are useful in our further derivations due to the relation between the entropy and the MMSE unveiled in [21].[3] Next, we give a lemma concerning the positive semidefiniteness of a certain class of matrices closely related with MMSE estimation.

Consider the setting described in the introduction, $\boldsymbol{Y} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W}$. For a given realization of the observations vector $\boldsymbol{Y} = \boldsymbol{y}$, the MMSE estimator, $\widehat{\boldsymbol{X}}(\boldsymbol{y})$, is given by the conditional mean

$$\widehat{\boldsymbol{X}}(\boldsymbol{y}) = \mathsf{E}\left\{\boldsymbol{X} | \boldsymbol{Y} = \boldsymbol{y}\right\}.$$

We now define the conditional MMSE matrix, $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$, as the mean-square error matrix conditioned on the fact that the received vector is equal to $\boldsymbol{Y} = \boldsymbol{y}$. Formally

$$\begin{aligned} \boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y}) &\triangleq \mathsf{E}\left\{(\boldsymbol{X} - \widehat{\boldsymbol{X}}(\boldsymbol{y}))(\boldsymbol{X} - \widehat{\boldsymbol{X}}(\boldsymbol{y}))^T \middle| \boldsymbol{Y} = \boldsymbol{y}\right\} \\ &= \mathsf{E}\left\{\boldsymbol{X}\boldsymbol{X}^T \middle| \boldsymbol{Y} = \boldsymbol{y}\right\} \\ &\quad - \mathsf{E}\left\{\boldsymbol{X} | \boldsymbol{Y} = \boldsymbol{y}\right\} \mathsf{E}\left\{\boldsymbol{X}^T \middle| \boldsymbol{Y} = \boldsymbol{y}\right\}. \end{aligned} \quad (11)$$

From this definition, it is clear that $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ is a positive semidefinite matrix.

Now, the MMSE matrix $\mathbf{E}_{\boldsymbol{X}}$ can be calculated by averaging $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ in (11) with respect to the distribution of vector $\boldsymbol{Y}$ as

$$\mathbf{E}_{\boldsymbol{X}} = \mathsf{E}\left\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}. \quad (12)$$

See below the last lemma in this section.

Lemma 9: For a given random vector $\boldsymbol{X} \in \mathbb{R}^n$, it follows that $\mathsf{E}\left\{\boldsymbol{X}\boldsymbol{X}^T\right\} \geq \mathsf{E}\left\{\boldsymbol{X}\right\} \mathsf{E}\left\{\boldsymbol{X}^T\right\}$.

Proof: Simply note that

$$\begin{aligned} &\mathsf{E}\left\{\boldsymbol{X}\boldsymbol{X}^T\right\} - \mathsf{E}\left\{\boldsymbol{X}\right\} \mathsf{E}\left\{\boldsymbol{X}^T\right\} \\ &\qquad = \mathsf{E}\left\{(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X}\right\})(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X}\right\})^T\right\} \geq \mathbf{0}, \end{aligned}$$

where last inequality follows from the fact that the expectation operator preserves positive semidefiniteness. ∎

[3] Strictly speaking the relation found in [21] concerns the quantities of mutual information and MMSE, but it is still useful for our problem because the entropy $h(\boldsymbol{Y})$ and the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ have the same dependence on $\boldsymbol{\Lambda}$ up to a constant additive term.

III. Main result of the paper

Once all the mathematical preliminaries have been presented, in this section we give the main result of the paper, namely, the concavity of the entropy power function $N(\boldsymbol{Y})$ in (6), with respect to the diagonal elements of $\boldsymbol{\Lambda}$. Prior to proving this result, we present a weaker result concerning the concavity of the entropy function $h(\boldsymbol{Y})$, which is key in proving the concavity of the entropy power.

A. Warm up: An entropy inequality

Theorem 1: Assume $\boldsymbol{Y} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W}$, where $\boldsymbol{X}$ is arbitrarily distributed and $\boldsymbol{W}$ follows a zero-mean white Gaussian distribution. Then the entropy $h(\boldsymbol{Y})$ is a concave function of the diagonal elements of $\boldsymbol{\Lambda}$, i.e.,

$$\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y}) \leq \mathbf{0}.$$

Furthermore, the entries of the Hessian matrix of the entropy function $h(\boldsymbol{Y})$ with respect to $\boldsymbol{\lambda}$ are given by

$$\begin{aligned} &\left[\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})\right]_{ij} \\ &= -\frac{1}{2}\mathsf{E}\left\{\left(\mathsf{E}\left\{X_i X_j | \boldsymbol{Y}\right\} - \mathsf{E}\left\{X_i | \boldsymbol{Y}\right\} \mathsf{E}\left\{X_j | \boldsymbol{Y}\right\}\right)^2\right\}, \end{aligned} \quad (13)$$

which can be written more compactly as

$$\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y}) = -\frac{1}{2}\mathsf{E}\left\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y}) \circ \boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}. \quad (14)$$

Proof: For the computations leading to (13) and (14) see Appendix I. Once the expression in (14) is obtained, concavity (or negative semidefiniteness of the Hessian matrix) follows straightforwardly taking into account that the matrix $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ defined in (11) is positive semidefinite $\forall \boldsymbol{y}$, Lemma 3, and from the fact that the expectation operator preserves the semidefiniteness. ∎

B. Multivariate extension of Costa's EPI

Theorem 2: Assume $\boldsymbol{Y} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W}$, where $\boldsymbol{X}$ is arbitrarily distributed and $\boldsymbol{W}$ follows a zero-mean white Gaussian distribution. Then the entropy power $N(\boldsymbol{Y})$ is a concave function of the diagonal elements of $\boldsymbol{\Lambda}$, i.e.,

$$\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) \leq \mathbf{0}.$$

Moreover, the Hessian matrix of the entropy power function $N(\boldsymbol{Y})$ with respect to $\boldsymbol{\lambda}$ is given by

$$\begin{aligned} &\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) \\ &\quad = \frac{N(\boldsymbol{Y})}{n}\left(\frac{\mathbf{d}_{\mathbf{E}_{\boldsymbol{X}}} \mathbf{d}_{\mathbf{E}_{\boldsymbol{X}}}^T}{n} - \mathsf{E}\left\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y}) \circ \boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\right), \end{aligned} \quad (15)$$

where we recall that $\mathbf{d}_{\mathbf{E}_{\boldsymbol{X}}}$ is a column vector with the diagonal entries of the matrix $\mathbf{E}_{\boldsymbol{X}}$ defined in (12).

Proof: First, let us prove (15). From the definition of the entropy power in (1) and applying the chain rule we obtain

$$\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) = \frac{2N(\boldsymbol{Y})}{n}\left(\frac{2\nabla_{\boldsymbol{\lambda}} h(\boldsymbol{Y}) \nabla^T_{\boldsymbol{\lambda}} h(\boldsymbol{Y})}{n} + \nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})\right).$$

Now, replacing $\nabla_{\boldsymbol{\lambda}} h(\boldsymbol{Y})$ by its expression from [21, Eq. (61)]

$$[\nabla_{\boldsymbol{\lambda}} h(\boldsymbol{Y})]_i = \frac{1}{2}[\mathbf{E}_{\boldsymbol{X}}]_{ii} = \frac{1}{2}\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ii}\right\},$$

and incorporating the expression for $\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})$ calculated in (14), the result in (15) follows.

Now that a explicit expression for the Hessian matrix has been obtained, we wish to prove that it is negative semidefinite. Note from (15) that, except for a positive factor, the Hessian matrix $\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ is the sum of a rank one positive semidefinite matrix and the Hessian matrix of the entropy, which is negative semidefinite according to Theorem 1. Consequently, the definiteness of $\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ is unknown a priori, and some further developments are needed to determine it, which is what we do next.

Consider a family of positive semidefinite matrices $\mathbf{A} \in \mathbb{S}^n_+$, characterized by a certain vector parameter $\mathbf{v}$, $\mathbf{A} = \mathbf{A}(\mathbf{v})$. Applying Proposition 8 to each matrix in this family, we obtain

$$\mathbf{A}(\mathbf{v}) \circ \mathbf{A}(\mathbf{v}) \geq \frac{\mathbf{d}_{\mathbf{A}(\mathbf{v})}\mathbf{d}^T_{\mathbf{A}(\mathbf{v})}}{n}. \tag{16}$$

Since (16) is true for all possible values of $\mathbf{v}$, we have

$$\mathsf{E}\left\{\mathbf{A}(\boldsymbol{V}) \circ \mathbf{A}(\boldsymbol{V})\right\} \geq \frac{\mathsf{E}\left\{\mathbf{d}_{\mathbf{A}(\boldsymbol{V})}\mathbf{d}^T_{\mathbf{A}(\boldsymbol{V})}\right\}}{n}, \tag{17}$$

where now the parameter $\mathbf{v}$ has been considered to be a random variable, $\boldsymbol{V}$. Note that the distribution of $\boldsymbol{V}$ is arbitrary and does not affect the validity of (17). From Lemma 9 we know that

$$\mathsf{E}\left\{\mathbf{d}_{\mathbf{A}(\boldsymbol{V})}\mathbf{d}^T_{\mathbf{A}(\boldsymbol{V})}\right\} \geq \mathsf{E}\left\{\mathbf{d}_{\mathbf{A}(\boldsymbol{V})}\right\}\mathsf{E}\left\{\mathbf{d}^T_{\mathbf{A}(\boldsymbol{V})}\right\},$$

from which it follows that

$$\mathsf{E}\left\{\mathbf{A}(\boldsymbol{V}) \circ \mathbf{A}(\boldsymbol{V})\right\} \geq \frac{\mathsf{E}\left\{\mathbf{d}_{\mathbf{A}(\boldsymbol{V})}\right\}\mathsf{E}\left\{\mathbf{d}^T_{\mathbf{A}(\boldsymbol{V})}\right\}}{n}.$$

Since the operators $\mathbf{d}_{\mathbf{A}}$ and expectation commute we finally obtain

$$\mathsf{E}\left\{\mathbf{A}(\boldsymbol{V}) \circ \mathbf{A}(\boldsymbol{V})\right\} \geq \frac{\mathbf{d}_{\mathsf{E}\{\mathbf{A}(\boldsymbol{V})\}}\mathbf{d}^T_{\mathsf{E}\{\mathbf{A}(\boldsymbol{V})\}}}{n}.$$

Identifying $\mathbf{A}(\boldsymbol{V})$ with the random covariance error matrix $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})$ and using (12) the result in the theorem follows as

$$\frac{\mathbf{d}_{\mathbf{E}_{\boldsymbol{X}}}\mathbf{d}^T_{\mathbf{E}_{\boldsymbol{X}}}}{n} - \mathsf{E}\left\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y}) \circ \boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\} \leq \mathbf{0},$$

and $N(\boldsymbol{Y}) \geq 0$. ∎

IV. Conclusion

In this paper we have proved that, for $\boldsymbol{Y} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W}$ the functions $N(\boldsymbol{Y})$ and $h(\boldsymbol{Y})$ are concave with respect to the diagonal entries of $\boldsymbol{\Lambda}$ and have also given explicit expressions for the elements of the Hessian matrices $\nabla^2_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ and $\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})$.

Besides its theoretical interest and inherent beauty, the importance of the results presented in this work lie mainly in their potential applications, such as, the calculation of the optimal power allocation to maximize the mutual information for a given non-Gaussian constellation as described in [14].

Appendix I
Calculation of $\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})$

In this section we are interested in the calculation of the elements of the Hessian matrix $\left[\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})\right]_{ij}$, which are defined by

$$\left[\nabla^2_{\boldsymbol{\lambda}} h(\boldsymbol{Y})\right]_{ij} = \frac{\partial^2 h(\boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W})}{\partial\lambda_i\partial\lambda_j}.$$

First of all, using the properties of differential entropy we write

$$h(\boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W}) = h(\boldsymbol{X} + \boldsymbol{\Lambda}^{-1/2}\boldsymbol{W}) + \frac{1}{2}\log|\boldsymbol{\Lambda}|,$$

and recalling that we work with natural logarithms we have

$$\frac{\partial^2 h(\boldsymbol{\Lambda}^{1/2}\boldsymbol{X} + \boldsymbol{W})}{\partial\lambda_i\partial\lambda_j} = \frac{\partial^2 h(\boldsymbol{X} + \boldsymbol{\Lambda}^{-1/2}\boldsymbol{W})}{\partial\lambda_i\partial\lambda_j} - \frac{\delta_{ij}}{2\lambda_i^2}. \tag{18}$$

We are now interested in expanding the first term in the right hand side of last equation, so we define the diagonal matrix $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}^{-1}$ and the random vector $\boldsymbol{Z} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{Y}$. Thus $[\boldsymbol{\Gamma}]_{ii} = \gamma_i = 1/\lambda_i$ and $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{\Lambda}^{-1/2}\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{W}$.

Applying the chain rule we obtain

$$\begin{aligned}\frac{\partial^2 h(\boldsymbol{X} + \boldsymbol{\Lambda}^{-1/2}\boldsymbol{W})}{\partial\lambda_i\partial\lambda_j} &= \frac{1}{\lambda_i^2\lambda_j^2}\frac{\partial^2 h(\boldsymbol{X} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i\partial\gamma_j}\bigg|_{\boldsymbol{\Gamma}=\boldsymbol{\Lambda}^{-1}} \\ &\quad + \frac{2\delta_{ij}}{\lambda_i^3}\frac{\partial h(\boldsymbol{X} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i}\bigg|_{\boldsymbol{\Gamma}=\boldsymbol{\Lambda}^{-1}}.\end{aligned} \tag{19}$$

The expressions for the two terms

$$\frac{\partial^2 h(\boldsymbol{X} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i\partial\gamma_j} \quad \text{and} \quad \frac{\partial h(\boldsymbol{X} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i}$$

are given in Appendix II, where we also sketch how they can be computed, for further details see [14]. Using these results, the right hand side of the expression in (19) can be rewritten as

$$\begin{aligned}&\frac{1}{\lambda_i^2\lambda_j^2}\left(-\frac{1}{2}\mathsf{E}\left\{\frac{\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}\right\}\right)^2}{\gamma_i^2\gamma_j^2}\right\}\right. \\ &\left.\quad -\frac{\delta_{ij}}{2\gamma_i^2} + \mathsf{E}\left\{\frac{\mathsf{E}\left\{X_i^2|\boldsymbol{Z}\right\} - \left(\mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2}{\gamma_i^3}\right\}\delta_{ij}\right)\Bigg|_{\boldsymbol{\Gamma}=\boldsymbol{\Lambda}^{-1}} \\ &\quad + \frac{2\delta_{ij}}{\lambda_i^3}\left(\frac{1}{2\gamma_i^2}\left(\gamma_i - \mathsf{E}\left\{\left(X_i - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2\right\}\right)\right)\Bigg|_{\boldsymbol{\Gamma}=\boldsymbol{\Lambda}^{-1}}.\end{aligned}$$

Simplifying terms we obtain

$$\begin{aligned}&-\frac{1}{2}\mathsf{E}\left\{\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}\right\}\right)^2\right\} \\ &\quad - \frac{\delta_{ij}}{2\lambda_i^2} + \frac{\delta_{ij}}{\lambda_i}\mathsf{E}\left\{\mathsf{E}\left\{X_i^2|\boldsymbol{Z}\right\} - \left(\mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2\right\} \\ &\quad + \frac{\delta_{ij}}{\lambda_i^2} - \frac{\delta_{ij}}{\lambda_i}\mathsf{E}\left\{\left(X_i - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2\right\}.\end{aligned} \tag{20}$$

Finally, noting that

$$\mathsf{E}\left\{\left(X_i - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2\right\} = \mathsf{E}\left\{\mathsf{E}\left\{X_i^2|\boldsymbol{Z}\right\} - \left(\mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\right)^2\right\}$$
$$\mathsf{E}\left\{f(\boldsymbol{X})|\boldsymbol{Z}\right\} = \mathbb{E}\left\{f(\boldsymbol{X})|\boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}\right\} = \mathsf{E}\left\{f(\boldsymbol{X})|\boldsymbol{Y}\right\},$$

and plugging (20) in (18) we obtain the desired result in (13):

$$\frac{\partial^2 h(\mathbf{\Lambda}^{1/2}\boldsymbol{X}+\boldsymbol{W})}{\partial\lambda_i\partial\lambda_j} = -\frac{1}{2}\mathsf{E}\left\{\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Y}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Y}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Y}\right\}\right)^2\right\}.$$

By simple inspection of the entries of the Hessian matrix above, the result in (14) can be found.

Appendix II
Gradient and Hessian of $h(\boldsymbol{Z} = \boldsymbol{X} + \mathbf{\Gamma}^{1/2}\boldsymbol{W})$

The elements of the gradient of $h(\boldsymbol{Z} = \boldsymbol{X} + \mathbf{\Gamma}^{1/2}\boldsymbol{W})$ with respect to the diagonal elements of $\mathbf{\Gamma}$ can be found thanks to the complex multivariate de Bruijn's identity found in [22, Th. 4] adapted to the real case

$$\frac{\partial h(\boldsymbol{X}+\mathbf{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i} = \frac{1}{2}\mathsf{E}\left\{\left(\frac{\partial\log P_{\boldsymbol{Z}}(\boldsymbol{Z})}{\partial z_i}\right)^2\right\}. \quad (21)$$

The elements of the Hessian matrix can be found quite directly from the expressions found in [7, Eq. (50)] and in Villani's Lemma in [9] for the single dimensional second derivative $\mathrm{d}^2 h(\boldsymbol{X}+\sqrt{t}\boldsymbol{W})/\,\mathrm{d}t^2$ (see [14] for further details on the specific generalization to the multidimensional case):

$$\frac{\partial^2 h(\boldsymbol{X}+\mathbf{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i\partial\gamma_j} = -\frac{1}{2}\mathsf{E}\left\{\left(\frac{\partial^2\log P_{\boldsymbol{Z}}(\boldsymbol{Z})}{\partial z_i\partial z_j}\right)^2\right\}. \quad (22)$$

To further elaborate the expressions in (21) and (22) we see that we need to compute the gradient and Hessian of the function $\log P_{\boldsymbol{Z}}(\mathbf{z})$. The expression for the gradient has already been given in [21, Eq. (56)], [22, Eq. (105)]

$$\frac{\partial\log P_{\boldsymbol{Z}}(\boldsymbol{z})}{\partial z_i} = \frac{\mathsf{E}\left\{X_i|\boldsymbol{Z}=\boldsymbol{z}\right\} - z_i}{\gamma_i}. \quad (23)$$

The expression for the Hessian of $\log P_{\boldsymbol{Z}}(\mathbf{z})$ requires slightly more elaboration and here we only give a sketch, more details can be found in [14].

Differentiating (23) with respect to z_j we obtain

$$\frac{\partial^2\log P_{\boldsymbol{Z}}(\boldsymbol{z})}{\partial z_i\partial z_j} = \frac{1}{\gamma_i\gamma_j}\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}=\boldsymbol{z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}=\boldsymbol{z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}=\boldsymbol{z}\right\}\right) - \frac{\delta_{ij}}{\gamma_i}, \quad (24)$$

where we have used that [14]

$$\frac{\partial\mathsf{E}\left\{X_i|\boldsymbol{Z}=\boldsymbol{z}\right\}}{\partial z_j} = \frac{1}{\gamma_j}\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}=\boldsymbol{z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}=\boldsymbol{z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}=\boldsymbol{z}\right\}\right).$$

Plugging (23) into (21) and operating according to the derivation in [22, Eq. (106)] we obtain

$$\frac{\partial h(\boldsymbol{X}+\mathbf{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i} = \frac{1}{2\gamma_i^2}\left(\gamma_i - \mathsf{E}\left\{(X_i - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\})^2\right\}\right).$$

Similarly, plugging (24) into (22) we obtain the desired expression for the Hessian as

$$\frac{\partial^2 h(\boldsymbol{X}+\mathbf{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i\partial\gamma_j} = -\frac{1}{2}\mathsf{E}\left\{\left(\frac{\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}\right\}}{\gamma_i\gamma_j} - \frac{\delta_{ij}}{\gamma_i}\right)^2\right\},$$

which can be expanded as

$$\frac{\partial^2 h(\boldsymbol{X}+\mathbf{\Gamma}^{1/2}\boldsymbol{W})}{\partial\gamma_i\partial\gamma_j} = -\frac{1}{2}\mathsf{E}\left\{\frac{\left(\mathsf{E}\left\{X_iX_j|\boldsymbol{Z}\right\} - \mathsf{E}\left\{X_i|\boldsymbol{Z}\right\}\mathsf{E}\left\{X_j|\boldsymbol{Z}\right\}\right)^2}{\gamma_i^2\gamma_j^2}\right\} - \frac{\delta_{ij}}{2\gamma_i^2} + \mathsf{E}\left\{\frac{\mathsf{E}\left\{X_i^2|\boldsymbol{Z}\right\} - (\mathsf{E}\left\{X_i|\boldsymbol{Z}\right\})^2}{\gamma_i^3}\right\}\delta_{ij}.$$

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst Tech. J*, vol. 27, no. 3, pp. 379–423, 1948.

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. and Control*, vol. 2, pp. 101–112, 1959.

[3] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. on Information Theory*, vol. 11, no. 2, pp. 267–271, 1965.

[4] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. on Information Theory*, vol. 52, no. 5, pp. 2165–2166, 2006.

[5] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. on Information Theory*, vol. 20, no. 2, pp. 279–280, 1974.

[6] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. on Information Theory*, vol. 51, no. 7, pp. 2577–2593, 2005.

[7] M. H. M. Costa, "A new entropy power inequality," *IEEE Trans. on Information Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[8] A. Dembo, "Simple proof of the concavity of the entropy power with respect to added Gaussian noise," *IEEE Trans. on Information Theory*, vol. 35, no. 4, pp. 887–888, 1989.

[9] C. Villani, "A short proof of the "concavity of entropy power"," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1695–1696, 2000.

[10] D. Guo, S. Shamai, and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE International Symposium on Information Theory (ISIT'06)*, July 2006.

[11] O. Rioul, "Information theoretic proofs of entropy power inequalities," *arXiv:0704.1751v1 [cs.IT]*, 2007.

[12] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. on Information Theory*, vol. 31, no. 5, pp. 607–615, 1985.

[13] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[14] M. Payaró and D. P. Palomar, "Hessian matrix and concavity properties of mutual information and entropy in linear vector Gaussian channels," *Submitted to IEEE Trans. on Information Theory*, 2008.

[15] R. Bhatia, *Positive definite matrices*. Princeton University Press, 2007.

[16] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985.

[17] ——, *Topics on matrix analysis*. Cambridge University Press, 1991.

[18] D. S. Bernstein, *Matrix mathematics*. Princeton University Press, 2005.

[19] C. R. Johnson and H. M. Shapiro, "Mathematical aspects of the relative gain array $A\circ A^{-T}$," *SIAM Journal on Algebraic and Discrete Methods*, vol. 7, p. 627, 1986.

[20] G. P. H. Styan, "Hadamard products and multivariate statistical analysis," *Linear Algebra Appl.*, vol. 6, pp. 217–240, 1973.

[21] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[22] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. on Information Theory*, vol. 52, no. 1, pp. 141–154, 2006.